Unigram mixtures and the EM algorithm

Guillaume Obozinski

INRIA - Ecole Normale Supérieure - Paris

RussIR summer school

Yaroslavl, August 6-10th 2012

The bag-of-word model, a vector-space representation of documents

Given

- a vocabulary of size d,

Represent a document consisting of N words

$$(w_1, \dots, w_N)$$

The bag-of-word model, a vector-space representation of documents

Given

- a vocabulary of size d,

Represent a document consisting of N words

$$(w_1, \ldots, w_N)$$

as x the vector of counts, or the vector of frequencies of the number of appearances of each of the words (possibly corrected with *tf-idf*):

$$x = \begin{bmatrix} x_1 \\ \vdots \\ x_d \end{bmatrix} \in \mathbb{N}_+^d, \quad \text{or } [0,1]_+^d, \quad \text{or } \mathbb{R}^d.$$

The bag-of-word model, a vector-space representation of documents

Given

- a vocabulary of size d,

Represent a document consisting of N words

$$(w_1, \ldots, w_N)$$

as x the vector of counts, or the vector of frequencies of the number of appearances of each of the words (possibly corrected with *tf-idf*):

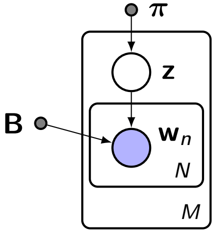
$$x = \begin{bmatrix} x_1 \\ \vdots \\ x_d \end{bmatrix} \in \mathbb{N}_+^d, \quad \text{or } [0,1]_+^d, \quad \text{or } \mathbb{R}^d.$$

Document collection

$$X = \begin{bmatrix} | & & | \\ x^{(1)} & \ldots & x^{(M)} \\ | & & | \end{bmatrix} = \begin{bmatrix} x_1^{(1)} & & x_1^{(M)} \\ \vdots & \ddots & \vdots \\ x_d^{(1)} & & x_d^{(M)} \end{bmatrix} \in \mathbb{R}^{d \times M}$$

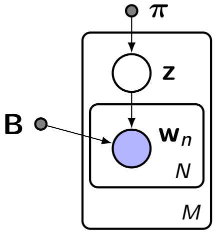
Multinomial mixture model (Unigram mixture)

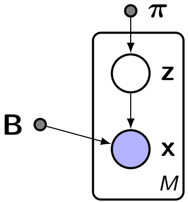
- K topics
- $\mathbf{z}$ component indicator vector
- $\mathbf{z} = (z_1, \dots, z_K)^\top \in \{0,1\}^K$
- $\mathbf{z} \sim \mathcal{M}(1, (\pi_1, \dots, \pi_K))$
- $p(\mathbf{z}) = \prod_{k=1}^{K} \pi_k^{z_k}$



Multinomial mixture model (Unigram mixture)

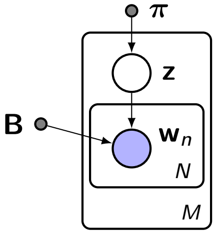
- K topics
- $\mathbf{z}$ component indicator vector
- $\mathbf{z} = (z_1, \ldots, z_K)^\top \in \{0,1\}^K$
- $\mathbf{z} \sim \mathcal{M}(1, (\pi_1, \ldots, \pi_K))$
- $p(\mathbf{z}) = \prod_{k=1}^{K} \pi_k^{z_k}$

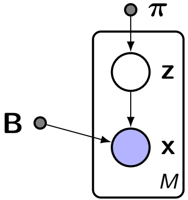




Multinomial mixture model (Unigram mixture)

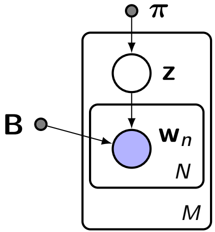
- K topics
- $\mathbf{z}$ component indicator vector
- $\mathbf{z} = (z_1, \ldots, z_K)^\top \in \{0,1\}^K$
- $\mathbf{z} \sim \mathcal{M}(1, (\pi_1, \ldots, \pi_K))$
- $p(\mathbf{z}) = \prod_{k=1}^{K} \pi_k^{z_k}$
- $\mathbf{w}_n \mid \{z_k = 1\} \sim \mathcal{M}(1, (b_{1k}, \ldots, b_{dk}))$
- $p(w_{nj} = 1 \mid z_k = 1) = b_{jk}$

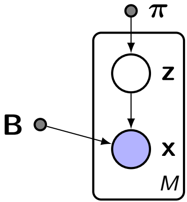




Multinomial mixture model (Unigram mixture)

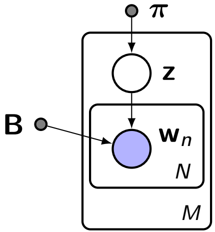
- K topics
- $\mathbf{z}$ component indicator vector
- $\mathbf{z} = (z_1, \dots, z_K)^\top \in \{0,1\}^K$
- $\mathbf{z} \sim \mathcal{M}(1, (\pi_1, \dots, \pi_K))$
- $p(\mathbf{z}) = \prod_{k=1}^{K} \pi_k^{z_k}$
- $\mathbf{w}_n \mid \{z_k = 1\} \sim \mathcal{M}(1, (b_{1k}, \dots, b_{dk}))$
- $p(w_{nj} = 1 \mid z_k = 1) = b_{jk}$
- $$p(\mathbf{w}, \mathbf{z}) = \prod_{n=1}^{N} \prod_{j=1}^{d} \prod_{k=1}^{K} \left(b_{jk}\pi_k\right)^{w_{nj} z_k}$$

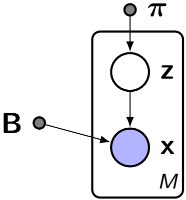




Multinomial mixture model (Unigram mixture)

- K topics
- $\mathbf{z}$ component indicator vector
- $\mathbf{z} = (z_1, \ldots, z_K)^\top \in \{0,1\}^K$
- $\mathbf{z} \sim \mathcal{M}(1, (\pi_1, \ldots, \pi_K))$
- $p(\mathbf{z}) = \prod_{k=1}^{K} \pi_k^{z_k}$
- $\mathbf{w}_n \mid \{z_k = 1\} \sim \mathcal{M}(1, (b_{1k}, \ldots, b_{dk}))$
- $p(w_{nj} = 1 \mid z_k = 1) = b_{jk}$
- $$p(\mathbf{w}, \mathbf{z}) = \prod_{n=1}^{N} \prod_{j=1}^{d} \prod_{k=1}^{K} \left(b_{jk}\pi_k\right)^{w_{nj} z_k}$$
- $$p(\mathbf{x}, \mathbf{z}) \propto \prod_{j=1}^{d} \prod_{k=1}^{K} \left(b_{jk}\pi_k\right)^{x_j z_k}$$





Applying maximum likelihood to the multinomial mixture

Let $\mathcal{Z} = \{z \in \{0,1\}^K \mid \sum_{k=1}^K z_k = 1\}$

Applying maximum likelihood to the multinomial mixture

Let $\mathcal{Z} = \{z \in \{0,1\}^K \mid \sum_{k=1}^K z_k = 1\}$

$$p(\mathbf{x}) =$$

Applying maximum likelihood to the multinomial mixture

Let $\mathcal{Z} = \{z \in \{0,1\}^K \mid \sum_{k=1}^K z_k = 1\}$

$$p(\mathbf{x}) = \sum_{\mathbf{z} \in \mathcal{Z}} p(\mathbf{x}, \mathbf{z})$$

Applying maximum likelihood to the multinomial mixture

Let $\mathcal{Z} = \{z \in \{0,1\}^K \mid \sum_{k=1}^K z_k = 1\}$

$$p(\mathbf{x}) = \sum_{\mathbf{z} \in \mathcal{Z}} p(\mathbf{x}, \mathbf{z}) = \sum_{\mathbf{z} \in \mathcal{Z}} \prod_{k=1}^{K} \Big[\prod_{j=1}^{d} b_{jk}^{x_j z_k} \Big] \pi_k^{z_k} =$$

Applying maximum likelihood to the multinomial mixture

Let $\mathcal{Z} = \{z \in \{0,1\}^K \mid \sum_{k=1}^K z_k = 1\}$

$$p(\mathbf{x}) = \sum_{\mathbf{z}\in\mathcal{Z}} p(\mathbf{x},\mathbf{z}) = \sum_{\mathbf{z}\in\mathcal{Z}} \prod_{k=1}^K \left[\prod_{j=1}^d b_{jk}^{x_j z_k}\right] \pi_k^{z_k} = \sum_{k=1}^K \left[\prod_{j=1}^d b_{jk}^{x_j}\right] \pi_k$$

Applying maximum likelihood to the multinomial mixture

Let $\mathcal{Z} = \{z \in \{0,1\}^K \mid \sum_{k=1}^K z_k = 1\}$

$$p(\mathbf{x}) = \sum_{\mathbf{z} \in \mathcal{Z}} p(\mathbf{x}, \mathbf{z}) = \sum_{\mathbf{z} \in \mathcal{Z}} \prod_{k=1}^{K} \left[\prod_{j=1}^{d} b_{jk}^{x_j z_k} \right] \pi_k^{z_k} = \sum_{k=1}^{K} \left[\prod_{j=1}^{d} b_{jk}^{x_j} \right] \pi_k$$

Issue

- The marginal log-likelihood $\ell(\mathbf{B}, \boldsymbol{\pi}) = \sum_i \log(p(\mathbf{x}^{(i)}))$ is now complicated

Applying maximum likelihood to the multinomial mixture

Let $\mathcal{Z} = \{z \in \{0,1\}^K \mid \sum_{k=1}^K z_k = 1\}$

$$p(\mathbf{x}) = \sum_{\mathbf{z}\in\mathcal{Z}} p(\mathbf{x},\mathbf{z}) = \sum_{\mathbf{z}\in\mathcal{Z}} \prod_{k=1}^{K} \left[\prod_{j=1}^{d} b_{jk}^{x_j z_k}\right] \pi_k^{z_k} = \sum_{k=1}^{K} \left[\prod_{j=1}^{d} b_{jk}^{x_j}\right] \pi_k$$

Issue

- The marginal log-likelihood $\ell(\mathbf{B}, \boldsymbol{\pi}) = \sum_i \log(p(\mathbf{x}^{(i)}))$ is now complicated
- No hope to find a simple solution to the maximum likelihood problem

Applying maximum likelihood to the multinomial mixture

Let $\mathcal{Z} = \{z \in \{0,1\}^K \mid \sum_{k=1}^K z_k = 1\}$

$$p(\mathbf{x}) = \sum_{\mathbf{z} \in \mathcal{Z}} p(\mathbf{x}, \mathbf{z}) = \sum_{\mathbf{z} \in \mathcal{Z}} \prod_{k=1}^{K} \left[\prod_{j=1}^{d} b_{jk}^{x_j z_k} \right] \pi_k^{z_k} = \sum_{k=1}^{K} \left[\prod_{j=1}^{d} b_{jk}^{x_j} \right] \pi_k$$

Issue

- The marginal log-likelihood $\ell(\mathbf{B}, \boldsymbol{\pi}) = \sum_i \log(p(\mathbf{x}^{(i)}))$ is now complicated
- No hope to find a simple solution to the maximum likelihood problem
- By contrast the complete log-likelihood has a rather simple form:

$$\tilde{\ell}(\mathbf{B}, \boldsymbol{\pi}) =$$

Applying maximum likelihood to the multinomial mixture

Let $\mathcal{Z} = \{z \in \{0,1\}^K \mid \sum_{k=1}^K z_k = 1\}$

$$p(\mathbf{x}) = \sum_{\mathbf{z}\in\mathcal{Z}} p(\mathbf{x},\mathbf{z}) = \sum_{\mathbf{z}\in\mathcal{Z}} \prod_{k=1}^K \left[\prod_{j=1}^d b_{jk}^{x_j z_k}\right] \pi_k^{z_k} = \sum_{k=1}^K \left[\prod_{j=1}^d b_{jk}^{x_j}\right] \pi_k$$

Issue

- The marginal log-likelihood $\ell(\mathbf{B},\boldsymbol{\pi}) = \sum_i \log(p(\mathbf{x}^{(i)}))$ is now complicated
- No hope to find a simple solution to the maximum likelihood problem
- By contrast the complete log-likelihood has a rather simple form:

$$\tilde{\ell}(\mathbf{B},\boldsymbol{\pi}) = \sum_{i=1}^M \log p(\mathbf{x}^{(i)},\mathbf{z}^{(i)})$$

Applying maximum likelihood to the multinomial mixture

Let $\mathcal{Z} = \{z \in \{0,1\}^K \mid \sum_{k=1}^K z_k = 1\}$

$$p(\mathbf{x}) = \sum_{\mathbf{z}\in\mathcal{Z}} p(\mathbf{x},\mathbf{z}) = \sum_{\mathbf{z}\in\mathcal{Z}} \prod_{k=1}^{K} \Big[\prod_{j=1}^{d} b_{jk}^{x_j z_k}\Big] \pi_k^{z_k} = \sum_{k=1}^{K} \Big[\prod_{j=1}^{d} b_{jk}^{x_j}\Big] \pi_k$$

Issue

- The marginal log-likelihood $\ell(\mathbf{B},\boldsymbol{\pi}) = \sum_i \log(p(\mathbf{x}^{(i)}))$ is now complicated
- No hope to find a simple solution to the maximum likelihood problem
- By contrast the complete log-likelihood has a rather simple form:

$$\tilde{\ell}(\mathbf{B},\boldsymbol{\pi}) = \sum_{i=1}^{M} \log p(\mathbf{x}^{(i)},\mathbf{z}^{(i)}) = \sum_{i,j,k} x_j^{(i)}\, z_k^{(i)} \log(b_{jk}) + \sum_{i,k} z_k^{(i)} \log(\pi_k)$$

Applying maximum likelihood to the multinomial mixture

$$\tilde{\ell}(\mathbf{B}, \boldsymbol{\pi}) = \sum_{i=1}^{M} \log p(\mathbf{x}^{(i)}, \mathbf{z}^{(i)}) = \sum_{i,j,k} x_j^{(i)} z_k^{(i)} \log(b_{jk}) + \sum_{i,k} z_k^{(i)} \log(\pi_k)$$

Applying maximum likelihood to the multinomial mixture

$$\tilde{\ell}(\mathbf{B}, \boldsymbol{\pi}) = \sum_{i=1}^{M} \log p(\mathbf{x}^{(i)}, \mathbf{z}^{(i)}) = \sum_{i,j,k} x_j^{(i)}\, z_k^{(i)} \log(b_{jk}) + \sum_{i,k} z_k^{(i)} \log(\pi_k)$$

- If we knew $\mathbf{z}^{(i)}$ we could maximize $\tilde{\ell}(\mathbf{B}, \boldsymbol{\pi})$.

Applying maximum likelihood to the multinomial mixture

$$\tilde{\ell}(\mathbf{B}, \boldsymbol{\pi}) = \sum_{i=1}^{M} \log p(\mathbf{x}^{(i)}, \mathbf{z}^{(i)}) = \sum_{i,j,k} x_j^{(i)} z_k^{(i)} \log(b_{jk}) + \sum_{i,k} z_k^{(i)} \log(\pi_k)$$

- If we knew $\mathbf{z}^{(i)}$ we could maximize $\tilde{\ell}(\mathbf{B}, \boldsymbol{\pi})$.
- If we knew $\mathbf{B}$ and $\boldsymbol{\pi}$, we could find the best $\mathbf{z}^{(i)}$ since we could compute the true a posteriori on $\mathbf{z}^{(i)}$ given $\mathbf{x}^{(i)}$:

Applying maximum likelihood to the multinomial mixture

$$\tilde{\ell}(\mathbf{B}, \boldsymbol{\pi}) = \sum_{i=1}^{M} \log p(\mathbf{x}^{(i)}, \mathbf{z}^{(i)}) = \sum_{i,j,k} x_j^{(i)} z_k^{(i)} \log(b_{jk}) + \sum_{i,k} z_k^{(i)} \log(\pi_k)$$

- If we knew $\mathbf{z}^{(i)}$ we could maximize $\tilde{\ell}(\mathbf{B}, \boldsymbol{\pi})$.
- If we knew $\mathbf{B}$ and $\boldsymbol{\pi}$, we could find the best $\mathbf{z}^{(i)}$ since we could compute the true a posteriori on $\mathbf{z}^{(i)}$ given $\mathbf{x}^{(i)}$:

$$p(z_k = 1 \mid \mathbf{x}; \mathbf{B}, \boldsymbol{\pi}) = \frac{\pi_k \prod_{j=1}^{d} b_{jk}^{x_j}}{\sum_{k'=1}^{K} \pi_{k'} \prod_{j=1}^{d} b_{jk'}^{x_j}}$$

Applying maximum likelihood to the multinomial mixture

$$\tilde{\ell}(\mathbf{B}, \boldsymbol{\pi}) = \sum_{i=1}^{M} \log p(\mathbf{x}^{(i)}, \mathbf{z}^{(i)}) = \sum_{i,j,k} x_j^{(i)}\, z_k^{(i)} \log(b_{jk}) + \sum_{i,k} z_k^{(i)} \log(\pi_k)$$

- If we knew $\mathbf{z}^{(i)}$ we could maximize $\tilde{\ell}(\mathbf{B}, \boldsymbol{\pi})$.
- If we knew $\mathbf{B}$ and $\boldsymbol{\pi}$, we could find the best $\mathbf{z}^{(i)}$ since we could compute the true a posteriori on $\mathbf{z}^{(i)}$ given $\mathbf{x}^{(i)}$:

$$p(z_k = 1 \mid \mathbf{x}; \mathbf{B}, \boldsymbol{\pi}) = \frac{\pi_k \prod_{j=1}^{d} b_{jk}^{x_j}}{\sum_{k'=1}^{K} \pi_{k'} \prod_{j=1}^{d} b_{jk'}^{x_j}}$$

$\rightarrow$ Seems a chicken and egg problem...

Applying maximum likelihood to the multinomial mixture

$$\tilde{\ell}(\mathbf{B}, \boldsymbol{\pi}) = \sum_{i=1}^{M} \log p(\mathbf{x}^{(i)}, \mathbf{z}^{(i)}) = \sum_{i,j,k} x_j^{(i)} z_k^{(i)} \log(b_{jk}) + \sum_{i,k} z_k^{(i)} \log(\pi_k)$$

- If we knew $\mathbf{z}^{(i)}$ we could maximize $\tilde{\ell}(\mathbf{B}, \boldsymbol{\pi})$.
- If we knew $\mathbf{B}$ and $\boldsymbol{\pi}$, we could find the best $\mathbf{z}^{(i)}$ since we could compute the true a posteriori on $\mathbf{z}^{(i)}$ given $\mathbf{x}^{(i)}$:

$$p(z_k = 1 \mid \mathbf{x}; \mathbf{B}, \boldsymbol{\pi}) = \frac{\pi_k \prod_{j=1}^{d} b_{jk}^{x_j}}{\sum_{k'=1}^{K} \pi_{k'} \prod_{j=1}^{d} b_{jk'}^{x_j}}$$

$\rightarrow$ Seems a chicken and egg problem...

- In addition, we want to solve

$$\max_{\mathbf{B}, \boldsymbol{\pi}} \sum_i \log \left(\sum_{\mathbf{z}^{(i)}} p(\mathbf{x}^{(i)}, \mathbf{z}^{(i)}) \right)$$

Applying maximum likelihood to the multinomial mixture

$$\tilde{\ell}(\mathbf{B}, \boldsymbol{\pi}) = \sum_{i=1}^{M} \log p(\mathbf{x}^{(i)}, \mathbf{z}^{(i)}) = \sum_{i,j,k} x_j^{(i)} \, z_k^{(i)} \log(b_{jk}) + \sum_{i,k} z_k^{(i)} \log(\pi_k)$$

- If we knew $\mathbf{z}^{(i)}$ we could maximize $\tilde{\ell}(\mathbf{B}, \boldsymbol{\pi})$.
- If we knew $\mathbf{B}$ and $\boldsymbol{\pi}$, we could find the best $\mathbf{z}^{(i)}$ since we could compute the true a posteriori on $\mathbf{z}^{(i)}$ given $\mathbf{x}^{(i)}$:

$$p(z_k = 1 \mid \mathbf{x}; \mathbf{B}, \boldsymbol{\pi}) = \frac{\pi_k \prod_{j=1}^{d} b_{jk}^{x_j}}{\sum_{k'=1}^{K} \pi_{k'} \prod_{j=1}^{d} b_{jk'}^{x_j}}$$

→ Seems a chicken and egg problem...

- In addition, we want to solve

$$\max_{\mathbf{B}, \boldsymbol{\pi}} \sum_i \log \Big(\sum_{\mathbf{z}^{(i)}} p(\mathbf{x}^{(i)}, \mathbf{z}^{(i)}) \Big) \quad \text{and not} \quad \max_{\substack{\mathbf{B}, \boldsymbol{\pi}, \\ \mathbf{z}^{(1)}, \ldots, \mathbf{z}^{(M)}}} \sum_i \log p(\mathbf{x}^{(i)}, \mathbf{z}^{(i)})$$

Applying maximum likelihood to the multinomial mixture

$$\tilde{\ell}(\mathbf{B}, \boldsymbol{\pi}) = \sum_{i=1}^{M} \log p(\mathbf{x}^{(i)}, \mathbf{z}^{(i)}) = \sum_{i,j,k} x_j^{(i)} \, z_k^{(i)} \log(b_{jk}) + \sum_{i,k} z_k^{(i)} \log(\pi_k)$$

- If we knew $\mathbf{z}^{(i)}$ we could maximize $\tilde{\ell}(\mathbf{B}, \boldsymbol{\pi})$.
- If we knew $\mathbf{B}$ and $\boldsymbol{\pi}$, we could find the best $\mathbf{z}^{(i)}$ since we could compute the true a posteriori on $\mathbf{z}^{(i)}$ given $\mathbf{x}^{(i)}$:

$$p(z_k = 1 \mid \mathbf{x}; \mathbf{B}, \boldsymbol{\pi}) = \frac{\pi_k \prod_{j=1}^{d} b_{jk}^{x_j}}{\sum_{k'=1}^{K} \pi_{k'} \prod_{j=1}^{d} b_{jk'}^{x_j}}$$

→ Seems a chicken and egg problem...

- In addition, we want to solve

$$\max_{\mathbf{B}, \boldsymbol{\pi}} \sum_i \log \Big(\sum_{\mathbf{z}^{(i)}} p(\mathbf{x}^{(i)}, \mathbf{z}^{(i)}) \Big) \quad \text{and not} \quad \max_{\substack{\mathbf{B}, \boldsymbol{\pi}, \\ \mathbf{z}^{(1)}, \ldots, \mathbf{z}^{(M)}}} \sum_i \log p(\mathbf{x}^{(i)}, \mathbf{z}^{(i)})$$

- Can we still use the intuitions above to construct an algorithm maximizing the marginal likelihood?

Principle of the Expectation-Maximization Algorithm

$$\log p(\mathbf{x};\boldsymbol{\theta}) \quad =$$

Principle of the Expectation-Maximization Algorithm

$$\log p(\mathbf{x};\boldsymbol{\theta}) = \log \sum_{\mathbf{z}} p(\mathbf{x},\mathbf{z};\boldsymbol{\theta})$$

Principle of the Expectation-Maximization Algorithm

$$\log p(\mathbf{x};\boldsymbol{\theta}) = \log \sum_{\mathbf{z}} p(\mathbf{x},\mathbf{z};\boldsymbol{\theta}) = \log \sum_{\mathbf{z}} q(\mathbf{z}) \frac{p(\mathbf{x},\mathbf{z};\boldsymbol{\theta})}{q(\mathbf{z})}$$

Principle of the Expectation-Maximization Algorithm

$$\begin{aligned}\log p(\mathbf{x};\boldsymbol{\theta}) &= \log \sum_{\mathbf{z}} p(\mathbf{x},\mathbf{z};\boldsymbol{\theta}) = \log \sum_{\mathbf{z}} q(\mathbf{z}) \frac{p(\mathbf{x},\mathbf{z};\boldsymbol{\theta})}{q(\mathbf{z})} \\ &\geq \sum_{\mathbf{z}} q(\mathbf{z}) \log \frac{p(\mathbf{x},\mathbf{z};\boldsymbol{\theta})}{q(\mathbf{z})}\end{aligned}$$

Principle of the Expectation-Maximization Algorithm

$$\begin{aligned}\log p(\mathbf{x};\boldsymbol{\theta}) &= \log \sum_{\mathbf{z}} p(\mathbf{x},\mathbf{z};\boldsymbol{\theta}) = \log \sum_{\mathbf{z}} q(\mathbf{z}) \frac{p(\mathbf{x},\mathbf{z};\boldsymbol{\theta})}{q(\mathbf{z})} \\ &\geq \sum_{\mathbf{z}} q(\mathbf{z}) \log \frac{p(\mathbf{x},\mathbf{z};\boldsymbol{\theta})}{q(\mathbf{z})} \\ &\quad = \mathbb{E}_q[\log p(\mathbf{x},\mathbf{z};\boldsymbol{\theta})] + H(q)\end{aligned}$$

Principle of the Expectation-Maximization Algorithm

$$\begin{aligned}
\log p(\mathbf{x};\boldsymbol{\theta}) &= \log \sum_{\mathbf{z}} p(\mathbf{x},\mathbf{z};\boldsymbol{\theta}) = \log \sum_{\mathbf{z}} q(\mathbf{z}) \frac{p(\mathbf{x},\mathbf{z};\boldsymbol{\theta})}{q(\mathbf{z})} \\
&\geq \sum_{\mathbf{z}} q(\mathbf{z}) \log \frac{p(\mathbf{x},\mathbf{z};\boldsymbol{\theta})}{q(\mathbf{z})} \\
&= \mathbb{E}_q[\log p(\mathbf{x},\mathbf{z};\boldsymbol{\theta})] + H(q) =: \mathcal{L}(q,\boldsymbol{\theta})
\end{aligned}$$

Principle of the Expectation-Maximization Algorithm

$$
\begin{aligned}
\log p(\mathbf{x};\boldsymbol{\theta}) &= \log \sum_{\mathbf{z}} p(\mathbf{x},\mathbf{z};\boldsymbol{\theta}) = \log \sum_{\mathbf{z}} q(\mathbf{z}) \frac{p(\mathbf{x},\mathbf{z};\boldsymbol{\theta})}{q(\mathbf{z})} \\
&\geq \sum_{\mathbf{z}} q(\mathbf{z}) \log \frac{p(\mathbf{x},\mathbf{z};\boldsymbol{\theta})}{q(\mathbf{z})} \\
&= \mathbb{E}_q[\log p(\mathbf{x},\mathbf{z};\boldsymbol{\theta})] + H(q) =: \mathcal{L}(q,\boldsymbol{\theta})
\end{aligned}
$$

- This shows that $\mathcal{L}(q,\boldsymbol{\theta}) \leq \log p(\mathbf{x};\boldsymbol{\theta})$

Principle of the Expectation-Maximization Algorithm

$$
\begin{aligned}
\log p(\mathbf{x};\boldsymbol{\theta}) &= \log \sum_{\mathbf{z}} p(\mathbf{x},\mathbf{z};\boldsymbol{\theta}) = \log \sum_{\mathbf{z}} q(\mathbf{z}) \frac{p(\mathbf{x},\mathbf{z};\boldsymbol{\theta})}{q(\mathbf{z})} \\
&\geq \sum_{\mathbf{z}} q(\mathbf{z}) \log \frac{p(\mathbf{x},\mathbf{z};\boldsymbol{\theta})}{q(\mathbf{z})} \\
&= \mathbb{E}_q[\log p(\mathbf{x},\mathbf{z};\boldsymbol{\theta})] + H(q) =: \mathcal{L}(q,\boldsymbol{\theta})
\end{aligned}
$$

- This shows that $\mathcal{L}(q,\boldsymbol{\theta}) \leq \log p(\mathbf{x};\boldsymbol{\theta})$
- Moreover: $\boldsymbol{\theta} \mapsto \mathcal{L}(q,\boldsymbol{\theta})$ is a **concave** function.

Principle of the Expectation-Maximization Algorithm

$$
\begin{aligned}
\log p(\mathbf{x};\boldsymbol{\theta}) &= \log \sum_{\mathbf{z}} p(\mathbf{x},\mathbf{z};\boldsymbol{\theta}) = \log \sum_{\mathbf{z}} q(\mathbf{z}) \frac{p(\mathbf{x},\mathbf{z};\boldsymbol{\theta})}{q(\mathbf{z})} \\
&\geq \sum_{\mathbf{z}} q(\mathbf{z}) \log \frac{p(\mathbf{x},\mathbf{z};\boldsymbol{\theta})}{q(\mathbf{z})} \\
&= \mathbb{E}_q[\log p(\mathbf{x},\mathbf{z};\boldsymbol{\theta})] + H(q) =: \mathcal{L}(q,\boldsymbol{\theta})
\end{aligned}
$$

- This shows that $\mathcal{L}(q,\boldsymbol{\theta}) \leq \log p(\mathbf{x};\boldsymbol{\theta})$
- Moreover: $\boldsymbol{\theta} \mapsto \mathcal{L}(q,\boldsymbol{\theta})$ is a **concave** function.
- Finally it is possible to show that

$$\mathcal{L}(q,\boldsymbol{\theta}) = \log p(\mathbf{x};\boldsymbol{\theta}) - KL(q||p(\cdot|\mathbf{x};\boldsymbol{\theta}))$$

Principle of the Expectation-Maximization Algorithm

$$\begin{aligned}\log p(\mathbf{x};\boldsymbol{\theta}) &= \log \sum_{\mathbf{z}} p(\mathbf{x},\mathbf{z};\boldsymbol{\theta}) = \log \sum_{\mathbf{z}} q(\mathbf{z}) \frac{p(\mathbf{x},\mathbf{z};\boldsymbol{\theta})}{q(\mathbf{z})} \\ &\geq \sum_{\mathbf{z}} q(\mathbf{z}) \log \frac{p(\mathbf{x},\mathbf{z};\boldsymbol{\theta})}{q(\mathbf{z})} \\ &\quad = \mathbb{E}_q[\log p(\mathbf{x},\mathbf{z};\boldsymbol{\theta})] + H(q) =: \mathcal{L}(q,\boldsymbol{\theta})\end{aligned}$$

- This shows that $\mathcal{L}(q,\boldsymbol{\theta}) \leq \log p(\mathbf{x};\boldsymbol{\theta})$
- Moreover: $\boldsymbol{\theta} \mapsto \mathcal{L}(q,\boldsymbol{\theta})$ is a **concave** function.
- Finally it is possible to show that

$$\mathcal{L}(q,\boldsymbol{\theta}) = \log p(\mathbf{x};\boldsymbol{\theta}) - KL(q||p(\cdot|\mathbf{x};\boldsymbol{\theta}))$$

So that if we set $q(\mathbf{z}) = p(\mathbf{z} \mid \mathbf{x};\boldsymbol{\theta}^{(t)})$ then

$$L(q,\boldsymbol{\theta}^{(t)}) = p(\mathbf{x};\theta^{(t)}).$$

Principle of the Expectation-Maximization Algorithm

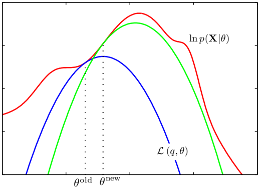
$$\begin{aligned}
\log p(\mathbf{x};\boldsymbol{\theta}) &= \log \sum_{\mathbf{z}} p(\mathbf{x},\mathbf{z};\boldsymbol{\theta}) = \log \sum_{\mathbf{z}} q(\mathbf{z}) \frac{p(\mathbf{x},\mathbf{z};\boldsymbol{\theta})}{q(\mathbf{z})} \\
&\geq \sum_{\mathbf{z}} q(\mathbf{z}) \log \frac{p(\mathbf{x},\mathbf{z};\boldsymbol{\theta})}{q(\mathbf{z})} \\
&= \mathbb{E}_q[\log p(\mathbf{x},\mathbf{z};\boldsymbol{\theta})] + H(q) =: \mathcal{L}(q,\boldsymbol{\theta})
\end{aligned}$$

- This shows that $\mathcal{L}(q,\boldsymbol{\theta}) \leq \log p(\mathbf{x};\boldsymbol{\theta})$
- Moreover: $\boldsymbol{\theta} \mapsto \mathcal{L}(q,\boldsymbol{\theta})$ is a **concave** function.
- Finally it is possible to show that

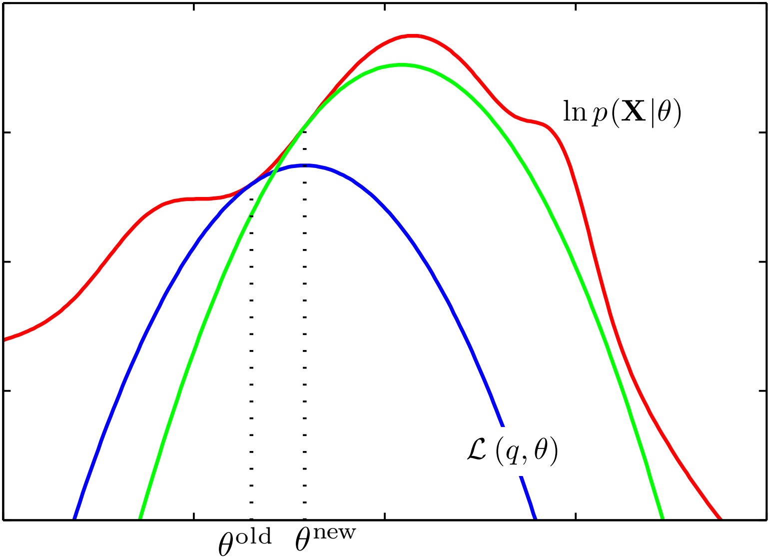
$$\mathcal{L}(q,\boldsymbol{\theta}) = \log p(\mathbf{x};\boldsymbol{\theta}) - KL(q||p(\cdot|\mathbf{x};\boldsymbol{\theta}))$$

So that if we set $q(\mathbf{z}) = p(\mathbf{z} \mid \mathbf{x};\boldsymbol{\theta}^{(t)})$ then

$$L(q,\boldsymbol{\theta}^{(t)}) = p(\mathbf{x};\theta^{(t)}).$$



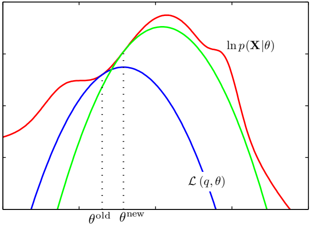
A graphical idea of the EM algorithm



Expectation Maximization algorithm

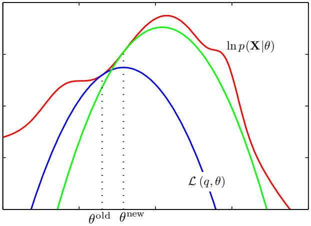
Expectation step

Maximization step



$$\boldsymbol{\theta}^{\text{old}} = \boldsymbol{\theta}^{(t-1)}$$
$$\boldsymbol{\theta}^{\text{new}} = \boldsymbol{\theta}^{(t)}$$

Expectation Maximization algorithm



Expectation step

1. $q(\mathbf{z}) = p(\mathbf{z} \mid \mathbf{x}; \boldsymbol{\theta}^{(t-1)})$

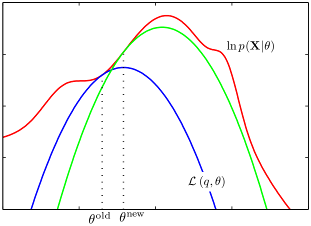
Maximization step

$$\boldsymbol{\theta}^{\text{old}} = \boldsymbol{\theta}^{(t-1)}$$
$$\boldsymbol{\theta}^{\text{new}} = \boldsymbol{\theta}^{(t)}$$

Expectation Maximization algorithm

Expectation step

1. $q(\mathbf{z}) = p(\mathbf{z} \mid \mathbf{x}; \boldsymbol{\theta}^{(t-1)})$
2. $$\mathcal{L}(q, \boldsymbol{\theta}) = \mathbb{E}_q \big[\log p(\mathbf{x}, \mathbf{z}; \boldsymbol{\theta}^{(t-1)}) \big] + H(q)$$



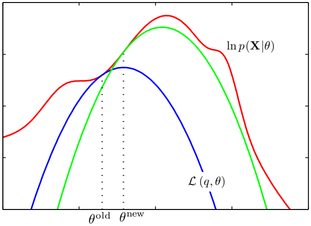
Maximization step

$$\boldsymbol{\theta}^{\text{old}} = \boldsymbol{\theta}^{(t-1)}$$
$$\boldsymbol{\theta}^{\text{new}} = \boldsymbol{\theta}^{(t)}$$

Expectation Maximization algorithm

Expectation step

1. $q(\mathbf{z}) = p(\mathbf{z} \mid \mathbf{x}; \boldsymbol{\theta}^{(t-1)})$
2. $$\mathcal{L}(q, \boldsymbol{\theta}) = \mathbb{E}_q \big[\log p(\mathbf{x}, \mathbf{z}; \boldsymbol{\theta}^{(t-1)}) \big] + H(q)$$



Maximization step

1. $\boldsymbol{\theta}^{(t)} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}} \, \mathbb{E}_q \big[\log p(\mathbf{x}, \mathbf{z}; \boldsymbol{\theta}^{(t-1)}) \big]$

$$\boldsymbol{\theta}^{\text{old}} = \boldsymbol{\theta}^{(t-1)}$$
$$\boldsymbol{\theta}^{\text{new}} = \boldsymbol{\theta}^{(t)}$$

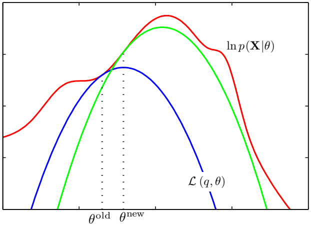
Expectation Maximization algorithm

Initialize $\boldsymbol{\theta} = \boldsymbol{\theta}_0$

WHILE (Not converged)

Expectation step

1. $q(\mathbf{z}) = p(\mathbf{z} \mid \mathbf{x}; \boldsymbol{\theta}^{(t-1)})$
2. $$\mathcal{L}(q, \boldsymbol{\theta}) = \mathbb{E}_q \big[\log p(\mathbf{x}, \mathbf{z}; \boldsymbol{\theta}^{(t-1)}) \big] + H(q)$$



Maximization step

1. $\boldsymbol{\theta}^{(t)} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}}\, \mathbb{E}_q \big[\log p(\mathbf{x}, \mathbf{z}; \boldsymbol{\theta}^{(t-1)}) \big]$

$$\boldsymbol{\theta}^{\text{old}} = \boldsymbol{\theta}^{(t-1)}$$
$$\boldsymbol{\theta}^{\text{new}} = \boldsymbol{\theta}^{(t)}$$

ENDWHILE

Expected complete log-likelihood

With the notation: $q_{ik}^{(t)} = \mathbb{P}_{q_i^{(t)}}(z_k^{(i)} = 1) = \mathbb{E}_{q_i^{(t)}}\left[z_k^{(i)}\right]$, we have

Expected complete log-likelihood

With the notation: $q_{ik}^{(t)} = \mathbb{P}_{q_i^{(t)}}(z_k^{(i)} = 1) = \mathbb{E}_{q_i^{(t)}}\left[z_k^{(i)}\right]$, we have

$$\mathbb{E}_{q^{(t)}}\left[\tilde{\ell}(\mathbf{B}, \boldsymbol{\pi})\right] \quad =$$

Expected complete log-likelihood

With the notation: $q_{ik}^{(t)} = \mathbb{P}_{q_i^{(t)}}(z_k^{(i)} = 1) = \mathbb{E}_{q_i^{(t)}}\left[z_k^{(i)}\right]$, we have

$$\mathbb{E}_{q^{(t)}}\left[\tilde{\ell}(\mathbf{B}, \boldsymbol{\pi})\right] = \mathbb{E}_{q^{(t)}}\left[\log p(\mathbf{X}, \mathbf{Z}; \mathbf{B}, \boldsymbol{\pi})\right]$$

Expected complete log-likelihood

With the notation: $q_{ik}^{(t)} = \mathbb{P}_{q_i^{(t)}}(z_k^{(i)} = 1) = \mathbb{E}_{q_i^{(t)}}\big[z_k^{(i)}\big]$, we have

$$\begin{aligned}\mathbb{E}_{q^{(t)}}\big[\tilde{\ell}(\mathbf{B}, \boldsymbol{\pi})\big] &= \mathbb{E}_{q^{(t)}}\big[\log p(\mathbf{X}, \mathbf{Z}; \mathbf{B}, \boldsymbol{\pi})\big] \\ &= \mathbb{E}_{q^{(t)}}\Big[\sum_{i=1}^{M} \log p(\mathbf{x}^{(i)}, \mathbf{z}^{(i)}; \mathbf{B}, \boldsymbol{\pi})\Big]\end{aligned}$$

Expected complete log-likelihood

With the notation: $q_{ik}^{(t)} = \mathbb{P}_{q_i^{(t)}}(z_k^{(i)} = 1) = \mathbb{E}_{q_i^{(t)}}\big[z_k^{(i)}\big]$, we have

$$\begin{aligned}
\mathbb{E}_{q^{(t)}}\big[\tilde{\ell}(\mathbf{B}, \boldsymbol{\pi})\big] &= \mathbb{E}_{q^{(t)}}\big[\log p(\mathbf{X}, \mathbf{Z}; \mathbf{B}, \boldsymbol{\pi})\big] \\
&= \mathbb{E}_{q^{(t)}}\Big[\sum_{i=1}^{M} \log p(\mathbf{x}^{(i)}, \mathbf{z}^{(i)}; \mathbf{B}, \boldsymbol{\pi})\Big] \\
&= \mathbb{E}_{q^{(t)}}\Big[\sum_{i,j,k} x_j^{(i)}\, z_k^{(i)} \log(b_{jk}) + \sum_{i,k} z_k^{(i)} \log(\pi_k)\Big]
\end{aligned}$$

Expected complete log-likelihood

With the notation: $q_{ik}^{(t)} = \mathbb{P}_{q_i^{(t)}}(z_k^{(i)} = 1) = \mathbb{E}_{q_i^{(t)}}\left[z_k^{(i)}\right]$, we have

$$\begin{aligned}
\mathbb{E}_{q^{(t)}}\left[\tilde{\ell}(\mathbf{B}, \boldsymbol{\pi})\right] &= \mathbb{E}_{q^{(t)}}\left[\log p(\mathbf{X}, \mathbf{Z}; \mathbf{B}, \boldsymbol{\pi})\right] \\
&= \mathbb{E}_{q^{(t)}}\left[\sum_{i=1}^{M} \log p(\mathbf{x}^{(i)}, \mathbf{z}^{(i)}; \mathbf{B}, \boldsymbol{\pi})\right] \\
&= \mathbb{E}_{q^{(t)}}\left[\sum_{i,j,k} x_j^{(i)}\, z_k^{(i)} \log(b_{jk}) + \sum_{i,k} z_k^{(i)} \log(\pi_k)\right] \\
&= \sum_{i,j,k} x_j^{(i)}\, \mathbb{E}_{q_i^{(t)}}\left[z_k^{(i)}\right] \log(b_{jk}) + \sum_{i,k} \mathbb{E}_{q_i^{(t)}}\left[z_k^{(i)}\right] \log(\pi_k)
\end{aligned}$$

Expected complete log-likelihood

With the notation: $q_{ik}^{(t)} = \mathbb{P}_{q_i^{(t)}}(z_k^{(i)} = 1) = \mathbb{E}_{q_i^{(t)}}\left[z_k^{(i)}\right]$, we have

$$\begin{aligned}
\mathbb{E}_{q^{(t)}}\left[\tilde{\ell}(\mathbf{B}, \boldsymbol{\pi})\right] &= \mathbb{E}_{q^{(t)}}\Big[\log p(\mathbf{X}, \mathbf{Z}; \mathbf{B}, \boldsymbol{\pi})\Big] \\
&= \mathbb{E}_{q^{(t)}}\left[\sum_{i=1}^{M} \log p(\mathbf{x}^{(i)}, \mathbf{z}^{(i)}; \mathbf{B}, \boldsymbol{\pi})\right] \\
&= \mathbb{E}_{q^{(t)}}\left[\sum_{i,j,k} x_j^{(i)} z_k^{(i)} \log(b_{jk}) + \sum_{i,k} z_k^{(i)} \log(\pi_k)\right] \\
&= \sum_{i,j,k} x_j^{(i)} \mathbb{E}_{q_i^{(t)}}\left[z_k^{(i)}\right] \log(b_{jk}) + \sum_{i,k} \mathbb{E}_{q_i^{(t)}}\left[z_k^{(i)}\right] \log(\pi_k) \\
&= \sum_{i,j,k} x_j^{(i)} q_{ik}^{(t)} \log(b_{jk}) + \sum_{i,k} q_{ik}^{(t)} \log(\pi_k)
\end{aligned}$$

Expectation step for the Multinomial mixture

We computed previously $q_i^{(t)}(\mathbf{z}^{(i)})$, which is a multinomial distribution defined by

$$q_i^{(t)}(\mathbf{z}^{(i)}) = p(\mathbf{z}^{(i)}|\mathbf{x}^{(i)}; \mathbf{B}^{(t-1)}, \boldsymbol{\pi}^{(t-1)})$$

Expectation step for the Multinomial mixture

We computed previously $q_i^{(t)}(\mathbf{z}^{(i)})$, which is a multinomial distribution defined by

$$q_i^{(t)}(\mathbf{z}^{(i)}) = p(\mathbf{z}^{(i)}|\mathbf{x}^{(i)}; \mathbf{B}^{(t-1)}, \boldsymbol{\pi}^{(t-1)})$$

Abusing notation we will denote $(q_{i1}^{(t)}, \ldots, q_{iK}^{(t)})$ the corresponding vector of probabilities defined by

$$q_{ik}^{(t)} = \mathbb{P}_{q_i^{(t)}}(z_k^{(i)} = 1) = \mathbb{E}_{q_i^{(t)}}\left[z_k^{(i)}\right]$$

Expectation step for the Multinomial mixture

We computed previously $q_i^{(t)}(\mathbf{z}^{(i)})$, which is a multinomial distribution defined by

$$q_i^{(t)}(\mathbf{z}^{(i)}) = p(\mathbf{z}^{(i)}|\mathbf{x}^{(i)}; \mathbf{B}^{(t-1)}, \boldsymbol{\pi}^{(t-1)})$$

Abusing notation we will denote $(q_{i1}^{(t)}, \ldots, q_{iK}^{(t)})$ the corresponding vector of probabilities defined by

$$q_{ik}^{(t)} = \mathbb{P}_{q_i^{(t)}}(z_k^{(i)} = 1) = \mathbb{E}_{q_i^{(t)}}\left[z_k^{(i)}\right]$$

$$q_{ik}^{(t)} = p(z_k^{(i)} = 1 \mid \mathbf{x}^{(i)}; \mathbf{B}^{(t-1)}, \boldsymbol{\pi}^{(t-1)}) = \frac{\pi_k^{(t-1)} \prod_{j=1}^d \left[b_{jk}^{(t-1)}\right]^{x_j^{(i)}}}{\sum_{k'=1}^K \pi_{k'}^{(t-1)} \prod_{j=1}^d \left[b_{jk'}^{(t-1)}\right]^{x_j^{(i)}}}$$

Maximization step for the Multinomial mixture

$$(\mathbf{B}^t, \boldsymbol{\pi}^t) = \underset{\mathbf{B},\boldsymbol{\pi}}{\operatorname{argmax}} \, \mathbb{E}_{q^{(t)}} \left[\tilde{\ell}(\mathbf{B}, \boldsymbol{\pi}) \right]$$

Maximization step for the Multinomial mixture

$$(\mathbf{B}^t, \boldsymbol{\pi}^t) = \underset{\mathbf{B},\boldsymbol{\pi}}{\operatorname{argmax}}\ \mathbb{E}_{q^{(t)}}\left[\tilde{\ell}(\mathbf{B}, \boldsymbol{\pi})\right]$$

This yields the updates:

$$\boxed{b_{jk}^{(t)} = \frac{\sum_i x_j^{(i)}\, q_{ik}^{(t)}}{\sum_i q_{ik}^{(t)}}} \quad \text{and} \quad \boxed{\pi_k^{(t)} = \frac{\sum_i q_{ik}^{(t)}}{\sum_{i,k'} q_{ik'}^{(t)}}}$$

Final EM algorithm for the Multinomial mixture model

Initialize $\boldsymbol{\theta} = \boldsymbol{\theta}_0$

WHILE (Not converged)

Expectation step

$$q_{ik}^{(t)} \leftarrow \frac{\pi_k^{(t-1)} \prod_{j=1}^d \left[b_{jk}^{(t-1)}\right]^{x_j^{(i)}}}{\sum_{k'=1}^K \pi_{k'}^{(t-1)} \prod_{j=1}^d \left[b_{jk'}^{(t-1)}\right]^{x_j^{(i)}}}$$

Maximization step

$$b_{jk}^{(t)} \leftarrow \frac{\sum_i x_j^{(i)} q_{ik}^{(t)}}{\sum_i q_{ik}^{(t)}} \qquad \text{and} \qquad \pi_k^{(t)} \leftarrow \frac{\sum_i q_{ik}^{(t)}}{\sum_i N^{(i)}}$$

ENDWHILE

The EM algorithm for the Gaussian mixture model

Gaussian mixture model

- K components
- $\mathbf{z}$ component indicator
- $\mathbf{z} = (z_1, \ldots, z_K)^\top \in \{0,1\}^K$
- $\mathbf{z} \sim \mathcal{M}(1, (\pi_1, \ldots, \pi_K))$
- $p(\mathbf{z}) = \prod_{k=1}^{K} \pi_k^{z_k}$

Gaussian mixture model

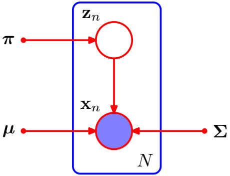
- K components
- $\mathbf{z}$ component indicator
- $\mathbf{z} = (z_1, \dots, z_K)^\top \in \{0,1\}^K$
- $\mathbf{z} \sim \mathcal{M}(1, (\pi_1, \dots, \pi_K))$
- $p(\mathbf{z}) = \prod_{k=1}^{K} \pi_k^{z_k}$
- $p(\mathbf{x}|\mathbf{z}; (\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)_k) = \sum_{k=1}^{K} z_k \, \mathcal{N}(\mathbf{x}; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$

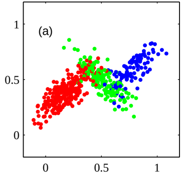
Gaussian mixture model

- K components
- $\mathbf{z}$ component indicator
- $\mathbf{z} = (z_1, \ldots, z_K)^\top \in \{0,1\}^K$
- $\mathbf{z} \sim \mathcal{M}(1, (\pi_1, \ldots, \pi_K))$
- $p(\mathbf{z}) = \prod_{k=1}^{K} \pi_k^{z_k}$
- $p(\mathbf{x}|\mathbf{z}; (\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)_k) = \sum_{k=1}^{K} z_k \, \mathcal{N}(\mathbf{x}; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$
- $p(\mathbf{x}) = \sum_{k=1}^{K} \pi_k \, \mathcal{N}(\mathbf{x}; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$

Gaussian mixture model

- K components
- $\mathbf{z}$ component indicator
- $\mathbf{z} = (z_1, \ldots, z_K)^\top \in \{0,1\}^K$
- $\mathbf{z} \sim \mathcal{M}(1, (\pi_1, \ldots, \pi_K))$
- $p(\mathbf{z}) = \prod_{k=1}^{K} \pi_k^{z_k}$
- $p(\mathbf{x}|\mathbf{z}; (\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)_k) = \sum_{k=1}^{K} z_k \mathcal{N}(\mathbf{x}; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$
- $p(\mathbf{x}) = \sum_{k=1}^{K} \pi_k \mathcal{N}(\mathbf{x}; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$
- Estimation: $\underset{\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k}{\operatorname{argmax}} \log \left[\sum_{k=1}^{K} \pi_k \mathcal{N}(\mathbf{x}; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \right]$

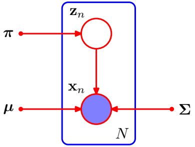




EM Algorithm for the Gaussian mixture model

Soit $\boldsymbol{\theta}^t = (\pi^t, (\boldsymbol{\mu}_k^t, \boldsymbol{\Sigma}_k^t)_k)$.

$$\prod_{i=1}^{n} p(\mathbf{z}^i, \mathbf{x}^i; \boldsymbol{\theta}) = \prod_{i=1}^{n} \prod_{k=1}^{K} \pi_k^{z_k^i} \Big(\mathcal{N}(x^i; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)\Big)^{z_k^i}$$



E step:

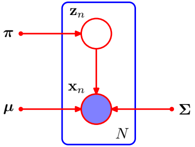
$p(\mathbf{z}^1, \ldots, \mathbf{z}^n | \mathbf{x}^1, \ldots, \mathbf{x}^n; \boldsymbol{\theta}^t) = \prod_{i=1}^{n} p(\mathbf{z}^i | \mathbf{x}^i; \boldsymbol{\theta}^t)$

$$q_k^i = P(z_k^i = 1 | x^i; \boldsymbol{\theta}^t) = \frac{p(x^i | z_k^i = 1; \boldsymbol{\theta}^t)\, P(z_k^i = 1; \boldsymbol{\theta}^t)}{p(x^i; \boldsymbol{\theta}^t)} =$$

EM Algorithm for the Gaussian mixture model

Soit $\boldsymbol{\theta}^t = (\pi^t, (\boldsymbol{\mu}_k^t, \boldsymbol{\Sigma}_k^t)_k)$.

$$\prod_{i=1}^{n} p(\mathbf{z}^i, \mathbf{x}^i; \boldsymbol{\theta}) = \prod_{i=1}^{n} \prod_{k=1}^{K} \pi_k^{z_k^i} \Big(\mathcal{N}(x^i; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \Big)^{z_k^i}$$



E step:

$$p(\mathbf{z}^1, \ldots, \mathbf{z}^n | \mathbf{x}^1, \ldots, \mathbf{x}^n; \boldsymbol{\theta}^t) = \prod_{i=1}^{n} p(\mathbf{z}^i | \mathbf{x}^i; \boldsymbol{\theta}^t)$$

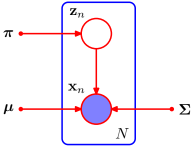
$$q_k^i = P(z_k^i = 1 | x^i; \boldsymbol{\theta}^t) = \frac{p(x^i | z_k^i = 1; \boldsymbol{\theta}^t)\, P(z_k^i = 1; \boldsymbol{\theta}^t)}{p(x^i; \boldsymbol{\theta}^t)} = \frac{\pi_k^t\, \mathcal{N}(x^i; \boldsymbol{\mu}_k^t, \boldsymbol{\Sigma}_k^t)}{\sum_\ell \pi_\ell^t\, \mathcal{N}(x^i; \boldsymbol{\mu}_\ell^t, \boldsymbol{\Sigma}_\ell^t)}$$

$$\mathbb{E}_q[\log p(\mathbf{z}, \mathbf{x} | \boldsymbol{\theta})] = \mathbb{E}_q \Big[\sum_{i,k} z_k^i \left(\log \pi_k + \log \mathcal{N}(x^i; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \right) \Big]$$

EM Algorithm for the Gaussian mixture model

Soit $\boldsymbol{\theta}^t = (\pi^t, (\boldsymbol{\mu}_k^t, \boldsymbol{\Sigma}_k^t)_k)$.

$$\prod_{i=1}^{n} p(\mathbf{z}^i, \mathbf{x}^i; \boldsymbol{\theta}) = \prod_{i=1}^{n} \prod_{k=1}^{K} \pi_k^{z_k^i} \Big(\mathcal{N}(x^i; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \Big)^{z_k^i}$$



E step:

$$p(\mathbf{z}^1, \ldots, \mathbf{z}^n | \mathbf{x}^1, \ldots, \mathbf{x}^n; \boldsymbol{\theta}^t) = \prod_{i=1}^{n} p(\mathbf{z}^i | \mathbf{x}^i; \boldsymbol{\theta}^t)$$

$$q_k^i = P(z_k^i = 1 | x^i; \boldsymbol{\theta}^t) = \frac{p(x^i | z_k^i = 1; \boldsymbol{\theta}^t)\, P(z_k^i = 1; \boldsymbol{\theta}^t)}{p(x^i; \boldsymbol{\theta}^t)} = \frac{\pi_k^t\, \mathcal{N}(x^i; \boldsymbol{\mu}_k^t, \boldsymbol{\Sigma}_k^t)}{\sum_\ell \pi_\ell^t\, \mathcal{N}(x^i; \boldsymbol{\mu}_\ell^t, \boldsymbol{\Sigma}_\ell^t)}$$

$$\mathbb{E}_q[\log p(\mathbf{z}, \mathbf{x} | \boldsymbol{\theta})] = \mathbb{E}_q\Big[\sum_{i,k} z_k^i \left(\log \pi_k + \log \mathcal{N}(x^i; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)\right) \Big]$$

$$= \sum_{i,k} q_k^i \log \pi_k - \frac{1}{2} q_k^i (x_i - \boldsymbol{\mu}_k)^\top \boldsymbol{\Sigma}_k^{-1} (x_i - \boldsymbol{\mu}_k) - \frac{1}{2} q_k^i \log((2\pi)^d |\boldsymbol{\Sigma}_k|)$$

EM Algorithm for the Gaussian mixture model II

$$Q(\boldsymbol{\theta}, \boldsymbol{\theta}^t) = \sum_{i,k} q_k^i \log \pi_k - \frac{1}{2} q_k^i (x_i - \boldsymbol{\mu}_k)^\top \boldsymbol{\Sigma}_k^{-1} (x_i - \boldsymbol{\mu}_k) - \frac{1}{2} q_k^i \log((2\pi)^d |\boldsymbol{\Sigma}_k|)$$

EM Algorithm for the Gaussian mixture model II

$$Q(\boldsymbol{\theta}, \boldsymbol{\theta}^t) = \sum_{i,k} q_k^i \log \pi_k - \frac{1}{2} q_k^i (x_i - \boldsymbol{\mu}_k)^\top \boldsymbol{\Sigma}_k^{-1} (x_i - \boldsymbol{\mu}_k) - \frac{1}{2} q_k^i \log((2\pi)^d |\boldsymbol{\Sigma}_k|)$$

M step:

$$\max_{\pi, (\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)_k} Q\Big((\pi, (\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)_k), \boldsymbol{\theta}^t \Big) \qquad \text{s.t.} \qquad \sum_k \pi_k = 1$$

EM Algorithm for the Gaussian mixture model II

$$Q(\boldsymbol{\theta}, \boldsymbol{\theta}^t) = \sum_{i,k} q_k^i \log \pi_k - \frac{1}{2} q_k^i (x_i - \boldsymbol{\mu}_k)^\top \boldsymbol{\Sigma}_k^{-1} (x_i - \boldsymbol{\mu}_k) - \frac{1}{2} q_k^i \log((2\pi)^d |\boldsymbol{\Sigma}_k|)$$

M step:

$$\max_{\pi, (\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)_k} Q\Big((\pi, (\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)_k), \boldsymbol{\theta}^t \Big) \qquad \text{s.t.} \qquad \sum_k \pi_k = 1$$

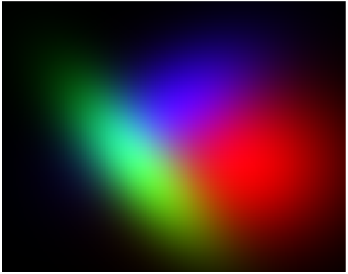
After calculations:

$$\boxed{n_k^{t+1} = \sum_i q_k^i} \qquad \boxed{\pi_k^{t+1} = \frac{n_k^{t+1}}{n}} \qquad \boxed{\boldsymbol{\mu}_k^{t+1} = \frac{1}{n_k^{t+1}} \sum_i q_k^i x_i}$$

$$\boxed{\boldsymbol{\Sigma}_k^{t+1} = \frac{1}{n_k^{t+1}} \sum_i q_k^i (x_i - \boldsymbol{\mu}_k^{t+1})(x_i - \boldsymbol{\mu}_k^{t+1})^\top}$$

EM Algorithm for the Gaussian mixture model III

$p(\mathbf{x}|\mathbf{z})$



$p(\mathbf{z}|\mathbf{x})$